View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by FORUM STATISTIKA DAN KOMPUTASI

Forum Statistika dan Komputasi, Oktober 2009 p : 14-18 Vol 14 No.2
ISSN : 0853-8115

PENDUGAAN REGESI SEMIPARAMETRIK DENGAN PENDEKATAN MODEL CAMPURAN LINEAR

(*Estimation of Semiparametric Regression using Linear Mixed Model Approach*)

Anik Djuraidah
Departemen Statistika FMIPA Institut Pertanian Bogor

E-mail : anikdjuraidah@gmail.com

Abstrak

Hubungan fungsional antara respon dengan peubah penjelas pada regresi linear berganda berbentuk parametrik dengan metode pendugaan parameternya adalah metode kuadrat terkecil. Pada regresi semiparametrik, hubungan fungsional antara respon dengan peubah penjelas dapat berbentuk parametrik atau nonparametrik. Metode yang banyak digunakan untuk pendugaan regresi semiparametrik adalah algoritma backfitting yang dikemukakan oleh Hastie & Tibshirani (1990). Pada penelitian ini pendugaan regresi parametrik didekati dengan model campuran linear. Keuntungan utama pendekatan dengan model campuran linear adalah menggunakan metode ML atau REML sehingga memberi kemudahan dalam seleksi model dan penarikan kesimpulan

Kata Kunci : Regresi semiparametrik, model aditif, model campuran linear, spline,

PENDAHULUAN

Pada regresi nonparametrik bentuk hubungan antara respon y dengan peubah penjelas x dinyatakan dalam model $y_i = s(x_i) + \varepsilon_i$ untuk $i = 1, \cdots, n$, dengan $s(.)$ adalah bentuk hubungan fungsional nonparametrik dan ε_i adalah galat acak. Stone (1985) memperluas model ini untuk peubah penjelas lebih dari satu, seperti pada regresi berganda, dan dikenal dengan model aditif. Untuk data $\{(y_i, x_{i1}, \cdots, x_{ip}),\ i = 1, \cdots, n\}$, model aditif didefinisikan sebagai

$$y_i = s_0 + \sum_{j=1}^{p} s_j(x_{ij}) + \varepsilon_i \qquad (1)$$

dengan $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$ adalah galat acak yang bebas stokastik terhadap peubah penjelas $\mathbf{x}_j = (x_{1j}, \cdots, x_{nj})'$ dan memenuhi $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, sedangkan s_j adalah bentuk hubungan fungsional antara respon dengan peubah penjelas $\mathbf{x}_j$.

Hubungan fungsional antara peubah penjelas dengan respon pada persamaan (1) dapat berbentuk parametrik, nonparametrik, atau gabungan keduanya. Bila semua hubungan fungsional antara respon dengan peubah penjelas pada model (1) berbentuk parametrik maka model (1) merupakan regresi linear berganda. Dan bila pada model (1) mengandung bentuk hubungan fungsional parametrik dan nonparametrik maka model (1) dikenal dengan regresi semiparametrik.

Metode untuk pendugaan regresi linear berganda adalah metode kuadrat terkecil. Untuk pendugaan model semiparametrik, Hastie & Tibshirani (1990) mengemukakan algoritma *backfitting*. Algoritma ini telah diperluas pada sebaran keluarga eksponensial dan dikenal sebagai model aditif terampat (*generalized additive model* selanjutnya disingkat GAM). Fungsi mulus yang paling banyak digunakan pada GAM adalah pemulus spline, di samping kernel dan fungsi penimbang lokal terkecil (*locally weigted* atau *lowess*). Meskipun GAM bersifat fleksibel dan efisien, akan tetapi algoritma *backfitting* dengan pemulus linear mempunyai kesulitan dalam seleksi model dan penarikan kesimpulan.

Eilers & Marx (1996), Ruppert & Carroll (1997) mengemukakan pemulus dimensi rendah yang disebut dengan P-spline atau regresi spline terpenalti (*penalized spline regression*). P-spline menggunakan jumlah basis spline yang sedikit dengan penalti kekasaran untuk mengontrol kemulusan dan mempunyai hubungan matematis yang sederhana dengan model campuran linear seperti yang dibahas oleh Fan dan Zhang (1998), Wang (1998), Brumback *et al* (1999), Vebyla *et al* (1999), French *et al* (2001), Kamman dan Wand

(2003), dan Wand (2003). Dalam model campuran linear, penduga P-spline adalah prediksi tak bias linear terbaik (*best linear unbiased linear prediction* selanjutnya disingkat BLUP) dan parameter pemulusnya merupakan ratio dari dua komponen ragam.

P-spline dapat dipandang sebagai model aditif untuk satu peubah penjelas. Kajian empiris pada model aditif satu peubah penjelas dengan P-spline dan pemulus spline menunjukkan bahwa perbedaan nilai parameter pemulus dan MSE (*mean square error*) antara keduanya relatif kecil (Djuraidah dan Aunuddin 2006). Model ini dapat dikembangkan untuk banyak peubah penjelas. Pada penelitian ini akan dikaji pendugaan regresi semiparametrik dengan pendekatan model campuran linear.

MODEL CAMPURAN LINEAR

Model linear campuran atau dikenal dengan komponen ragam merupakan perluasan dari model linear yaitu dengan menambahkan efek acak. Metode ini banyak digunakan dalam rancangan percobaan untuk data yang berkorelasi seperti pada percobaan pengukuran berulang (Searle *et al* 1992). Bentuk umum model linear campuran adalah

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

$$\text{dengan } \begin{bmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{bmatrix} \sim \left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}\right) \qquad (2)$$

dimana $\mathbf{X}$ adalah matriks desain dari efek tetap yang teramati, $\boldsymbol{\beta}$ adalah vektor parameter pengaruh efek tetap yang tidak diketahui, $\mathbf{Z}$ adalah matriks desain efek acak yang teramati, $\mathbf{u}$ adalah vektor efek acak yang tidak diketahui, dan $\boldsymbol{\varepsilon}$ adalah vektor galat acak yang tidak diketahui. Sehingga nilaitengah dan matriks ragam-peragam untuk $\mathbf{y}$ adalah $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ dan $\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$.

Penduga efek tetap dan efek acak ditentukan dari persamaan model campuran yaitu

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{u}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix}$$

Bila matriks $\mathbf{G}$ dan $\mathbf{R}$ diketahui, maka penduga bagi parameter $\boldsymbol{\beta}$ dan $\mathbf{u}$ adalah

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\\ \mathbf{Z}'\end{bmatrix}\mathbf{R}^{-1}\mathbf{y} \qquad (3)$$

Dari persamaan matriks di atas, penduga efek tetap $\boldsymbol{\beta}$ dapat dinyatakan sebagai

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

Penduga efek tetap $\hat{\boldsymbol{\beta}}$ adalah penduga GLS (*generalized least squares*) dan $\mathbf{X}\hat{\boldsymbol{\beta}}$ adalah penduga tak bias linear terbaik (*best linear unbiased linear prediction* selanjutnya disingkat BLUE) untuk $\mathbf{X}\boldsymbol{\beta}$.

Bila $\mathbf{y}$ mempunyai sebaran normal, maka $\hat{\boldsymbol{\beta}}$ adalah penduga kemungkinan maksimum (*maximum likelihood estimator* selanjutnya disingkat MLE). Penduga efek acak $\hat{\mathbf{u}}$ pada persamaan (3) dapat dinyatakan sebagai

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

yang merupakan BLUP (Searle *et al* 1992).

Metode yang paling banyak digunakan dalam pendugaan komponen ragam adalah metode kemungkinan maksimum (*maximum likelihood* selanjutnya disingkat ML) dan metode kemungkinan maksimum berkendala (*restricted maximum likelihood* selanjutnya disingkat REML). Metode REML menghasilkan penduga takbias bagi komponen ragam, sedangkan ML menghasilkan penduga yang bias.

REGRESI SPLINE TERPENALTI

Misalkan (x_i, y_i) adalah pengukuran pada peubah penjelas x dan peubah respon y untuk $1 \le i \le n$. Misalkan hubungan fungsional antara x dengan y dimodelkan sebagai

$$y_i = s(x_i) + \varepsilon_i \qquad (4)$$

dengan s adalah fungsi mulus, ε_i bebas stokastik dengan ragam σ^2. Model (4) adalah bentuk regresi nonparametrik yang paling sederhana dan banyak metode pendekatan yang dapat digunakan seperti yang dibahas oleh Eubank (1988), Green dan Silverman (1994), dan Simonoff (1996) Misalkan fungsi mulus s diduga dengan model regresi spline berderajat-p yaitu :

$$s(\mathbf{x};\boldsymbol{\beta}) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{k=1}^{K} u_{pk}(x-\kappa_k)_+^p \qquad (5)$$

dengan $\boldsymbol{\beta} = \left(\beta_0,\ldots\beta_p,u_{p1},\ldots,u_{pK}\right)'$ adalah vektor koefisien regresi spline, $p \ge 1$ adalah bilangan bulat positif, $(w)_+^p = w^p \mathbf{I}(w \ge 0)$ adalah basis fungsi pangkat terpotong berderajat p (*truncated power function* selanjutnya disingkat dengan FPT) dengan $\mathbf{I}$ fungsi indikator, dan $\kappa_1 < \ldots < \kappa_K$ adalah simpul tetap.

Penduga parameter $\hat{\boldsymbol{\beta}}$ ditentukan dengan minimisasi jumlah kuadrat terpenalti, yaitu J(s) yang didefinisikan sebagai :

$$J(s) = \sum_{i=1}^{n}(y_i - s(x_i;\boldsymbol{\beta}))^2 + \lambda\ \boldsymbol{\beta}'\mathbf{D}\ \boldsymbol{\beta} \qquad (6)$$

dengan λ adalah parameter pemulus, dan $\mathbf{D}$ adalah matriks semidefinit positif. Suku pertama pada J(s) adalah jumlah kuadrat galat dan suku keduanya adalah penalti kekasaran. Kriteria penentuan model pada persamaan (6) merupakan gabungan antara kriteria pada model regresi spline dengan kriteria dari pemulus spline. Sehingga minimisasi J(s) pada nilai λ tertentu akan memberikan kompromi antara kebaikan pengepasan dengan kehalusan kurva. Model aditif degan kriteria pendugaan ini disebut juga dengan regresi spline terpenalti.

Misalkan $\mathbf{T}$ adalah matriks desain untuk regresi spline dengan baris ke-i dari matriks $\mathbf{T}$ yaitu $\mathbf{T}_i = \left(1,\ x_i,\ \ldots\ x_i^p,\ (x_i-\kappa_1)_+^p,\ \ldots\ (x_i-\kappa_K)_+^p\right)$. Misalkan $\mathbf{D} = \text{diag}\ (\ \mathbf{0}_{p+1}, \mathbf{1}_K\ )$, maka dalam notasi matriks J(s) dapat dinyatakan sebagai

$$(\mathbf{y}-\mathbf{T}\ \boldsymbol{\beta})'(\mathbf{y}-\mathbf{T}\ \boldsymbol{\beta}) + \lambda\ \ \boldsymbol{\beta}'\mathbf{D}\ \boldsymbol{\beta} \qquad (7)$$

Minimisasi persamaan (7) menghasilkan penduga bagi parameter $\boldsymbol{\beta}$ yaitu

$$\hat{\boldsymbol{\beta}} = (\mathbf{T'T} + \lambda\ \mathbf{D})^{-1}\mathbf{T'y}$$

sehingga penduga regresi spline terpenalti adalah

$$\hat{\mathbf{y}} = \mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{T}(\mathbf{T'T} + \lambda\ \mathbf{D})^{-1}\mathbf{T'y} \qquad (8)$$

Regresi spline terpenalti dapat diformulasikan sebagai model campuran linear. Kunci hubungan antara regresi spline terpenalti dengan model campuran linear adalah bahwa penalti dari koefisien u_{pk} pada model (5) ekivalen dengan memperlakukan koefisien ini sebagai efek acak pada model campuran linear. Misalkan didefinisikan parameter

$$\boldsymbol{\beta}^* = (\beta_0, \beta_1, \cdots, \beta_p)', \quad \mathbf{u} = (u_{p1}, \ldots, u_{pK})',$$

dan matriks desain $\mathbf{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1 & \cdots & x_n^p \end{bmatrix}$, dan

$$\mathbf{Z} = \begin{bmatrix} (x_1-\kappa_1)_+^p & \cdots & (x_1-\kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_n-\kappa_1)_+^p & \cdots & (x_n-\kappa)_{K+}^{\ p} \end{bmatrix}$$

Kriteria spline terpenalti pada persamaan (6) jika dibagi dengan σ_ε^2 akan diperoleh

$$\frac{1}{\sigma_\varepsilon^2}\left(\ \mathbf{y}-\mathbf{X}\ \boldsymbol{\beta}^* - \mathbf{Z}\ \mathbf{u}\ \right)'\left(\ \mathbf{y}-\mathbf{X}\ \boldsymbol{\beta}^* - \mathbf{Z}\ \mathbf{u}\ \right) + \frac{\lambda}{\sigma_\varepsilon^2}\ \mathbf{u}'\mathbf{u} \qquad (9)$$

Persamaan (9) sama dengan kriteria BLUP dari model campuran linear dengan memperlakukan $\mathbf{u}$ sebagai koefisien dari efek acak pada`model campuran linear dengan $\text{cov}(\mathbf{u}) = \sigma_u^2\mathbf{I}$ dimana $\sigma_u^2 = \frac{\sigma_\varepsilon^2}{\lambda}$. Sehingga representasi regresi spline terpenalti dalam bentuk model campuran linear adalah

$$\mathbf{y} = \mathbf{X}\ \boldsymbol{\beta}^* + \mathbf{Z}\ \mathbf{u} + \boldsymbol{\varepsilon}\text{, dengan } \mathrm{E}\begin{bmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \mathbf{0} \text{ dan}$$

$$\text{cov}\begin{bmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\sigma_u^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \sigma_\varepsilon^2\mathbf{I}\end{bmatrix} \qquad (10)$$

BLUP untuk fungsi $s(\mathbf{x}) = (s(x_1), \ldots, s(x_n))'$ adalah

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}^* + \mathbf{Z}\hat{\mathbf{u}} \qquad (11)$$

dengan

$$\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}'\ \left(\ \sigma_u^2\ \mathbf{Z}\ \mathbf{Z}'\ +\ \sigma_\varepsilon^2\ \mathbf{I}\ \right)^{-1}\ \mathbf{X}\right)^{-1}\ \mathbf{X}'\ \left(\ \sigma_u^2\ \mathbf{Z}\ \mathbf{Z}'\ +\ \sigma_\varepsilon^2\ \mathbf{I}\ \right)^{-1}\mathbf{y}$$

dan $\quad \hat{\mathbf{u}} = \sigma_u^2\ \mathbf{Z}'\ \left(\ \sigma_u^2\ \mathbf{Z}\ \mathbf{Z}'\ +\ \sigma_\varepsilon^2\ \mathbf{I}\ \right)^{-1}\left(\mathbf{y} - \mathbf{X}\ \hat{\boldsymbol{\beta}}^*\right).$

Solusi $\hat{\mathbf{y}}$ di atas dapat dinyatakan dalam bentuk

$$\hat{\mathbf{y}} = \mathbf{C}'\left(\mathbf{C'C} + \lambda\ \mathbf{D}\right)^{-1}\mathbf{C'y} \qquad (12)$$

dengan $\mathbf{C} = [\mathbf{X}\quad \mathbf{Z}]$, $\mathbf{D} = \text{diag}(\ \mathbf{0}_{p+1}, \mathbf{1}_K\ )$ dan $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_u^2}$. Persamaan (12) ekivalen dengan solusi regresi spline terpenalti pada persamaan (5). Bukti ini menunjukkan bahwa BLUP bagi $s(\mathbf{x})$ ekivalen dengan penduga regresi spline terpenalti. Parameter pemulus λ merupakan rasio antara dua komponen ragam dari model campuran linear.

REGRESI SEMIPARAMETRIK

Misalkan regesi semiparametrik dengan 2 peubah penjelas X_1 dan X_2, dengan hubungan fungsional antara Y dengan X_1 berbentuk linear dan hubungan fungsional antara Y dengan X_2 berbentuk nonparametrik. Model ini dapat dinyatakan sebagai

$$y_i = \beta_0 + \beta_1 x_{1i} + s(x_{2i}) + \varepsilon_i \qquad (13)$$

dengan $s(x_{2i})$ adalah fungsi mulus untuk x_2. Fungsi mulus $s(x_2)$ dimodelkan sebagai regresi spline linear berderajat-1 maka model (13) ditulis sebagai

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sum_{k=1}^{K} u_k (x_i - \kappa_k)_+ + \varepsilon_i \qquad (14)$$

dengan $\kappa_1 < \ldots < \kappa_K$ adalah simpul tetap untuk peubah penjelas penjelas x_2, dan $(w)_+ = w\ \mathbf{I}(w \geq 0)$ adalah FPT berderajat-1 dengan $\mathbf{I}$ fungsi indikator. Model pada persamaan (14) diduga dengan minimisasi jumlah kuadrat terpenalti.

Misalkan vektor koefisien pada model (14) adalah $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2,\ u_1,\ \cdots,\ u_K)'$, penduga parameter $\hat{\boldsymbol{\beta}}$ ditentukan dengan minimisasi jumlah kuadrat terpenalti seperti pada persamaan (6) dan dipeoleh

$$\hat{\boldsymbol{\beta}} = (\mathbf{C'C} + \lambda\ \mathbf{D})^{-1}\mathbf{C'y}$$

sehingga penduga regresi semiparametrik adalah

$$\hat{\mathbf{y}} = \mathbf{C}(\mathbf{C'C} + \mathbf{D})^{-1}\mathbf{C'y}$$

dengan

$$\mathbf{C} = \begin{bmatrix} 1 & x_{11} & x_{21} & (x_{21}-\kappa_1)_+ & \cdots & (x_{21}-\kappa_K)_+ \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & (x_{2n}-\kappa_1)_+ & \cdots & (x_{2n}-\kappa_K)_+ \end{bmatrix}$$

dan $\mathbf{D} = \mathrm{diag}\left(0 \quad 0 \quad 0 \quad \lambda \mathbf{1}_{K\times 1}\right)$, serta λ adalah parameter pemulus peubah penjelas penjelas x_2 .

Formulasi regresi semiparametrik pada persamaan (13) ke dalam model linear campuran adalah dengan memperlakukan potongan polinomial u_k sebagai efek acak dalam model linear campuran. Jika didefinisikan

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)', \quad \mathbf{u} = (u_1 \quad \cdots \quad u_K)'$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} (x_{21} - \kappa_1)_+ & \cdots & (x_{21} - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_{2n} - \kappa_1)_+ & \cdots & (x_{2n} - \kappa_K)_+ \end{bmatrix}$$

maka penduga kuadrat terkecil terpenalti adalah ekivalen dengan BLUP model linear campuran

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \text{ dengan } \mathrm{E}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \mathbf{0} \text{ dan}$$

$$\mathrm{cov}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \sigma_2^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix}$$

Sehingga pendugaan model aditif dapat dilakukan dengan menggunakan model linear campuran. Penduga parameter pemulus untuk X_2 merupakan rasio antara dua komponen ragam yaitu

$$\lambda_2 = \frac{\sigma_\varepsilon^2}{\sigma_2^2}.$$

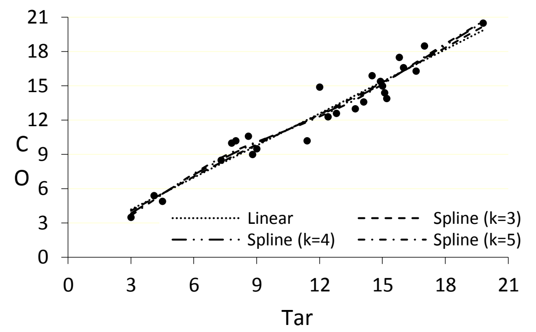
METODOLOGI PENELITIAN

Data yang digunakan pada penelitian ini berasal dari buku Mendenhall & Sincich (1992) yaitu tentang kandungan karbon monooksida /CO (mg), kandungan tar (mg) dan berat (g) dari 25 merek rokok. Kandungan CO sebagai peubah respon, sedangkan kandungan tar dan berat sebagai peubah penjelasnya.

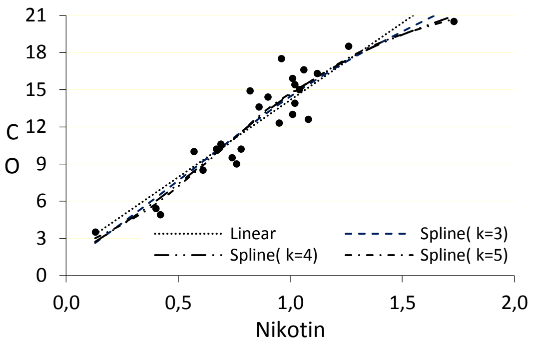
Pemodelan diawali dengan menentukan hubungan fungsional antara CO dengan tar dan berat. Pada peubah penjelas yang mempunyai hubungan nonparametrik dibangkitkan basis FPT dengan beberapa kemungkinan jumlah simpul yang sesuai dengan data. Dari dua peubah penjelas di atas dan jumlah simpul yang digunakan akan diperoleh beberapa alternatif model. Selanjutnya model semiparametrik diformulasikan dalam model campuran linear. Pendugaan parameter dan komponen ragam menggunakan metode REML dengan bantuan paket program SAS v9.1. Kriteria kebaikan model ditentukan dari nilai AIC dan plot galat model.

HASIL DAN PEMBAHASAN

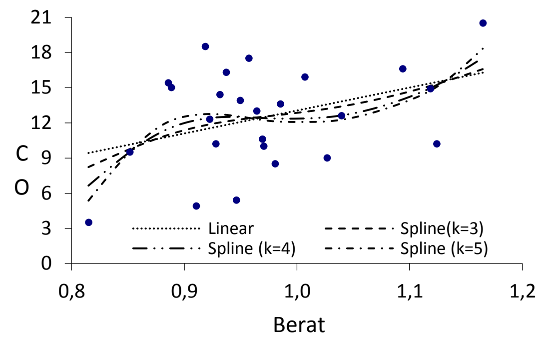
Koefisien korelasi antara kandungan tar dan kandungan nikotin sebesar 0.936, sehingga dalam model semiparametrik cukup dipilih salah satu dari kedua peubah penjelas tersebut. Pola hubungan fungsional antara kandungan CO dengan tar, nikotin, dan berat masing-masing disajikan pada Gambar 1, 2, dan 3. Pada Gambar 1 , tampak pola hubungan antara tar dengan CO adalah linear. Demikian juga dengan pola hubungan antara nikotin dengan CO pada Gambar 2 tampak linear. Sedangkan pola hubungan antara berat dengan CO tampak nonparametrik.



Gambar 1. Pola Hubungan Fungsional antara Tar dengan CO



Gambar 2. Pola Hubungan Fungsional antara Nikotin dengan CO



Gambar 3. Pola Hubungan Fungsional antara Berat dengan CO

Hasil pemodelan regresi linear dan regresi semiparametrik disajikan pada Tabel 1 dan Tabel 2. Nilai AIC pada model dengan peubah penjelas tar lebih kecil dibandingkan model dengan peubah penjelas nikotin. Pada Tabel 1 terlihat penambahan jumlah simpul pada peubah penjelas berat tidak menurunkan nilai AIC model. Model terbaik untuk CO mempunyai peubah penjelas tar dengan pola hubungan fungsional linear dan berat dengan hubungan fungsional spline derajat 2 dan jumlah simpul 3. Pola sisaan model terbaik yang disajikan pada Gambar 4 tampak tidak berpola.

Tabel 1. Nilai AIC model dengan peubah penjelas Tar dan Berat

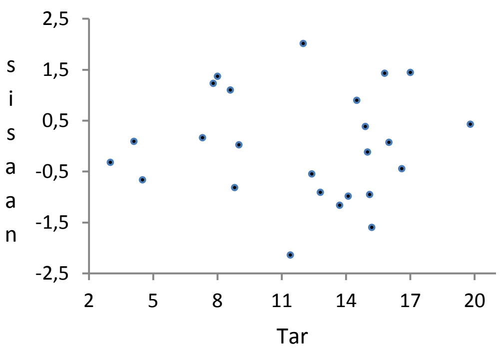
Hubungan Fungsional dengan CO		AIC
Tar	Berat	
Linear	Linear	76.4
Linear	Spline d=1, k=3	76.4
Linear	Spline d=1, k=4	76.4
Linear	Spline d=1, k=5	76.4
Linear	Spline d=2, k=3	66.0
Linear	Spline d=2, k=4	66.0
Linear	Spline d=2, k=5	66.0

d= derajat spline k= jumlah simpul

Tabel 2. Nilai AIC model dengan peubah penjelas Nikotin dan Berat

Hubungan Fungsional dengan CO		AIC
Nikotin	Berat	
Linear	Linear	91.6
Linear	Spline d=1, k=3	91.6
Linear	Spline d=1, k=4	91.6
Linear	Spline d=1, k=5	89.6
Linear	Spline d=2, k=3	81.8
Linear	Spline d=2, k=4	81.8
Linear	Spline d=2, k=5	81.8

d= derajat spline k= jumlah simpul



Gambar 4. Plot Sisaan Model Terbaik

KESIMPULAN

Nilai AIC model semiparametrik lebih kecil dibandingkan dengan model regresi linear. Peningkatan derajat polinomial spline menurunkan nilai AIC, meskipun penambahan jumlah simpul tidak menurunkan niali AIC. Pendugaan regresi semiparametrik dengan model linear campuran memberikan kemudahan dalam pendugaan model dengan komputasi yang cepat.

DAFTAR PUSTAKA

Brumback BA, Ruppert D, Wand MP. 1999. Comment on Variable selection and function estimation in additive nonparametric regression using a data-based prior by Shively, Kohn and Wood. *J Amer Stat Ass* 94: 794-797.

Djuraidah A & Aunuddin. 2006. Pendugaan Regresi Spline Terpenalti dengan Pendekatan Model Campuran. *Statistika* Jurnal Statistika FMIPA-UNISBA 6(1): 39-46.

Eilers PHC & Marx BD. 1996. Flexible smoothing with B-splines and penalties (with discussion). *Stat Sci* 11:89-121.

Fan J & Zhang JT. 1998. Comment on Smoothing spline models for the analysis of nested and crossed samples of curves by Brumback and Rice. *J Amer Stat Ass* 93: 961-994.

French JL, Kammann EE, Wand MP. 2001. Comment on Semiparametric nonlinear mixed-effects models and their applications by Ke and Wang. *J Amer Stat Ass* 96:1285-1288.

Hastie TJ & Tibshirani RJ. 1990. *Generalized Additive Models*. London: Chapman & Hall.

Kammann EE & Wand MP. 2003. Geoadditive models. *Appl Stat* 52:1-18.

Mendenhall W & Sincich T. 1992. Statistics for Engineering and the Sciences. 3rd ed. New York: Dellen Publishing

Ruppert D & Carroll RJ. 1997. Penalized regression splines. Unpublished manuscript. [terhubung berkala]. http://www.orie.cornell.edu/~davidr/papers/ index/index/index.html

Searle SR, Casella G, McCulloch CE. 1992. Variance Component. New York : John Wiley & Sons.

Stone CJ. 1985. Additive Regression and Other Nonparametric Models. *Ann Stat* 13: 689–705.

Verbyla AP, Cullis BR, Kenward MG, Welham SJ. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *J R Stat Soc*, Series C 48: 269-312.

Wand M. 2003. Smoothing and mixed models. *Comp Stat* 18:223–249.

Wang Y, 1998. Mixed effects smoothing spline analysis of variance. *J R Stat Soc*, Series B 60:159-174.